

# Reliability of Quadruplicated Serological Parameters in the Korean Genome and Epidemiology Study

Jae Jeong Yang<sup>1,2</sup>, Ji Hyun Yang<sup>1,2</sup>, Jimin Kim<sup>3</sup>, Lisa Y. Cho<sup>1,2</sup>, Boyoung Park<sup>1,2</sup>, Seung Hyun Ma<sup>1,2</sup>, Sang Hoon Song<sup>4</sup>, Won-Ki Min<sup>5</sup>, Sung Soo Kim<sup>3</sup>, Man Suck Park<sup>3</sup>, Sue K. Park<sup>1,2,6</sup>

<sup>1</sup>Department of Preventive Medicine, Seoul National University College of Medicine; <sup>2</sup>Cancer Research Institute, Seoul National University, Seoul; <sup>3</sup>Center for Genome Science, Korea National Institute of Health, Osong; <sup>4</sup>Department of Laboratory Medicine, Seoul National University College of Medicine; <sup>5</sup>Department of Laboratory Medicine, Asan Medical Center, University of Ulsan College of Medicine; <sup>6</sup>Department of Biomedical Science, Seoul National University Graduate School, Seoul, Korea

**OBJECTIVES:** The aim of this study was to evaluate whether clinical test values from different laboratories in the Korean Genome and Epidemiology Study (KoGES) can be integrated through a statistical adjustment algorithm with appropriate intra- and inter-laboratory reliability.

**METHODS:** External quality control data were obtained from the Korean Society for Laboratory Medicine and quadruplicated standardized serological samples (N = 3,200) were manufactured in order to check the intra- and inter-laboratory reliability for aspartic acid transaminase (AST), alanine transaminase (ALT), gamma-glutamyl transpeptidase ( $\gamma$ -GTP), blood urea nitrogen (BUN), creatinine, uric acid (UA), fasting blood sugar (FBS), cholesterol, and triglyceride (TG). As an index of inter- and intra-rater reliability, Pearson's correlation coefficient, intraclass correlation coefficients and kappa statistics were estimated. In addition, to detect the potential for data integration, we constructed statistical compensation models using linear regression analysis with residual analysis, and presented the R-square values.

**RESULTS:** All correlation coefficient values indicated good intra- and inter-laboratory reliability, which ranged from 0.842 to 1.000. Kappa coefficients were greater than 0.75 (0.75-1.00). All of the regression models based on the trial results had strong R-square values and zero sums of residuals. These results were consistent in the regression models using external quality control data.

**CONCLUSION:** The two laboratories in the KoGES have good intra- and inter-laboratory reliability for ten chemical test values, and data can be integrated through algorithmic statistical adjustment using regression equations.

**KEY WORDS:** Reliability, Statistical compensation model, Serological marker, The KoGES

## INTRODUCTION

Correspondence: Sue K. Park, MD, PhD

Department of Preventive Medicine, Seoul National University College of Medicine, 28 Yeongeong-dong, Jongno-gu, Seoul 110-799, Korea  
Tel: +82-2-740-8338 Fax: +82-2-747-4830 E-mail: suepark@snu.ac.kr

Received: Nov, 15 2010, Accepted: Apr 13, 2011, Published: May 19, 2011

This article is available from: <http://e-epih.org/>.

© 2011, Korean Society of Epidemiology

© This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Korean Genome and Epidemiology Study (KoGES), a large population-based genomic cohort study supported by the Korea Centers for Disease Control and Prevention (KCDC), investigates risk factors for major diseases among Koreans, in particular focusing on gene-environment and gene-gene interactions [1,2]. According to the standardized KoGES protocol 12 biochemical analyses (fasting blood sugar [FBS]; gamma-glutamyl transpeptidase [ $\gamma$ -GTP]; aspartic acid transaminase [AST]; alanine transaminase [ALT]; total cholesterol; triglyceride [TG]; high density lipoprotein [HDL]-cholesterol; low density lipoprotein [LDL]-cholesterol; albumin; blood urea nitrogen [BUN]; uric acid [UA]; and creatinine), complete blood cell

counts including eight hematologic indices (white blood cells [WBC]; red blood cells [RBC]; hemoglobin [Hb]; hematocrit [HCT]; mean corpuscular volume [MCV]; mean corpuscular hemoglobin [MCH]; mean corpuscular hemoglobin concentration [MCHC]; and platelets), and analysis of high sensitivity c-reactive protein (HS-CRP) was conducted at two clinical laboratories, and authenticated via external quality assessment (QA) by the Clinical and Laboratory Standards Institute (CLSI) and the Korean Society for Laboratory Medicine (KSLM).

Clinical results from multiple laboratories may differ systematically [3-5]. Despite standardization of the blood test in the KoGES, potential discrepancies that may lead to significantly biased results in the pooled analyses remain. Moreover, the fundamental differences between the laboratories involving instruments (Hitachi Co., Tokyo, Japan vs. Bayer HealthCare Ltd., Tarrytown, NY, USA), methods (enzymatic vs. colorimetry for  $\gamma$ -GTP) and reference values (adapted to each instrument) may result in heterogeneity.

To minimize the potential analytic drift, appropriate standardization such as calibration or statistical adjustment is warranted [6]. Thus, we investigated whether 1) the KoGES laboratories measure consistent test values regardless of time and environmental conditions, 2) the laboratories are able to produce the same test values for future data pooling, and 3) future pooled analysis for serological parameters is possible using a statistical compensation model even though absolute test values differ between laboratories. In the present study, the intra- and inter-laboratory reliability of the two clinical laboratories were assessed using quadruplicated standardized serological samples and external quality control (QC) data, and the possibility that serological parameters produced by different laboratories can

be integrated through statistical compensation models was evaluated.

## METHODS

To check the overall reliability of each laboratory, three years of QC data (2005, 2006, and 2007) for ten blood chemical tests (albumin, ALT, AST,  $\gamma$ -GTP, BUN, creatinine, UA, FBS, cholesterol, and TG) were obtained from the KSLM. To evaluate intra- and inter-laboratory reliability, quadruplicated standardized serological samples were prepared according to the Clinical and Laboratory Standards Institute (CLSI) guidelines [7]. Basic principles including 'at least 40 patient samples', 'at least 5 operating days', and 'analysis of each sample in duplicate within the same run' were adopted. The principle of 'at least 50% of samples outside the reference range' was revised based on the true range of each test result in the KoGES and the general distribution in the Korean population. Finally, 40 samples were divided into five groups with two outlier groups A and E (Table 1).

The standardized serological samples were manufactured at the Department of Laboratory Medicine at the Seoul National University College of Medicine. Eight anonymous patient samples were prepared for five consecutive operating days. During the repeated trial, a total of 3,200 test serological samples were manufactured (eight patients  $\times$  five operating days  $\times$  duplication  $\times$  repeated trial  $\times$  ten test items  $\times$  two laboratories). In detail, sample preparation was carried out as follows: 1) a nurse at the Department of Laboratory Medicine selected all serum samples with a volume of more than 3 mL before the day of the test, 2) a well-trained clinical pathologist selected eight serum samples

**Table 1.** Total number of standardized serological samples (N=40) prepared according to the test value range<sup>1</sup>

	Group A		Group B		Group C		Group D		Group E	
	Range of test value	Samples (%)	Range of test value	Samples (%)	Range of test value	Samples (%)	Range of test value	Samples (%)	Range of test value	Samples (%)
Albumin <sup>3</sup>	< 3	4 (10)	3-4	16 (40)	4-5	16 (40)	> 5	4 (10)	-	-
ALT	< 20	8 (20)	20-40	8 (20)	20-40	16 (40)	80-160	4 (10)	161-SL <sup>2</sup>	4 (10)
AST	< 20	8 (20)	20-40	12 (30)	20-40	12 (30)	80-160	4 (10)	161-SL <sup>2</sup>	4 (10)
$\gamma$ -GTP <sup>3</sup>	0-30	16 (40)	31-60	16 (40)	120-240	4 (10)	241-SL <sup>2</sup>	4 (10)	-	-
BUN	< 15	4 (10)	15-25	16 (40)	26-50	8 (20)	51-100	8 (20)	100-SL <sup>2</sup>	4 (10)
Creatinine	0-1.0	8 (20)	1.1-2.5	12 (30)	2.5-5.0	8 (20)	5-10	8 (20)	11-SL <sup>2</sup>	4 (10)
Uric acid	< 3.0	8 (20)	3-5	8 (20)	5-8	8 (20)	8-10	8 (20)	11-SL <sup>2</sup>	4 (10)
FBS	< 50	4 (10)	51-110	16 (40)	111-150	12 (30)	151-250	4 (10)	251-SL <sup>2</sup>	4 (10)
Cholesterol <sup>3</sup>	120-180	8 (20)	181-220	12 (30)	221-260	12 (30)	261-400	8 (20)	-	-
TG	< 75	4 (10)	75-125	12 (30)	125-200	12 (30)	200-300	8 (20)	301-SL <sup>2</sup>	4 (10)

ALT, alanine transaminase; AST, aspartic acid transaminase;  $\gamma$ -GTP, gamma-glutamyl transpeptidase; BUN, blood urea nitrogen; creatinine; UA, uric acid; FBS, fasting blood sugar; TG, triglyceride.

<sup>1</sup>Using 40 serum samples, the standardized serological samples were prepared from the number of specimens predetermined according to the range of test values. We prepared each sample twice, and thus 80 serum samples were acquired in this study; <sup>2</sup>The maximum value that could be detected by each instrument; <sup>3</sup>Albumin,  $\gamma$ -GTP, and cholesterol were divided into four categories based on both the KoGES data and normal ranges for the general Korean population.

on the morning of the test, 3) four 250  $\mu$ L aliquots of each serum (quadruplicated samples) were divided into plain tubes, and 4) a set of duplicated samples was delivered to each center. To minimize other sources of variation, sample delivery, storage conditions, and test time were consistent with a standardized protocol. On the day of the test, prepared samples were delivered to each laboratory between noon and 1 pm, and tests were conducted simultaneously at 3 pm. Serum selection and tests were not conducted on a Saturday, Sunday, or Monday. The first trial was conducted on November 29-30 and December 4-7, 2007. On the 6th day of the first trial, additional tests for FBS, BUN, UA, and albumin with unacceptable values outside the reference range (from Groups A or E) were conducted. The second trial was conducted over five days on December 11, 12, 13, 14, and 20, 2007.

The mean of the difference and the coefficient of variation (CV) between the duplicated test results at each laboratory were calculated to determine intra-rater reliability. Pearson's correlation coefficients and intra-class correlation (ICC) coefficients were estimated as indices of intra-rater reliability. To determine the inter-laboratory reliability, we estimated the mean and CV of the differences between the mean of 40 duplicated test values at Centers X and Y. The level of agreement between the two laboratories was assessed using Pearson's correlation coefficients, ICC coefficients and the Bland-Altman Plot [8]. Kappa statistics were used to measure the agreement in diagnosis (normal vs. abnormal). Coefficient values of 0.81-1.00 and  $>0.75$  indicated 'almost perfect' and 'substantial and excellent', respectively [9,10].

Statistical compensation models were constructed using linear regression analysis to determine the potential for data integration. The regression equation was as follows: (Test values at Center Y) =  $\alpha$  (intercept) +  $\beta$  (slope)  $\times$  (Test values at Center X). The model fitting was checked by both regression residual analysis and R-square values. All statistical analyses were performed using SAS version 9.2 (SAS Inc., Cary, NC, USA) and MedCalc version 11.5 (MedCalc Software, Inc, Mariakerke, Belgium).

The present study was approved by the Institutional Review Boards of Seoul National University Hospital and the National Cancer Center of Korea (C-0707-072-214).

## RESULTS

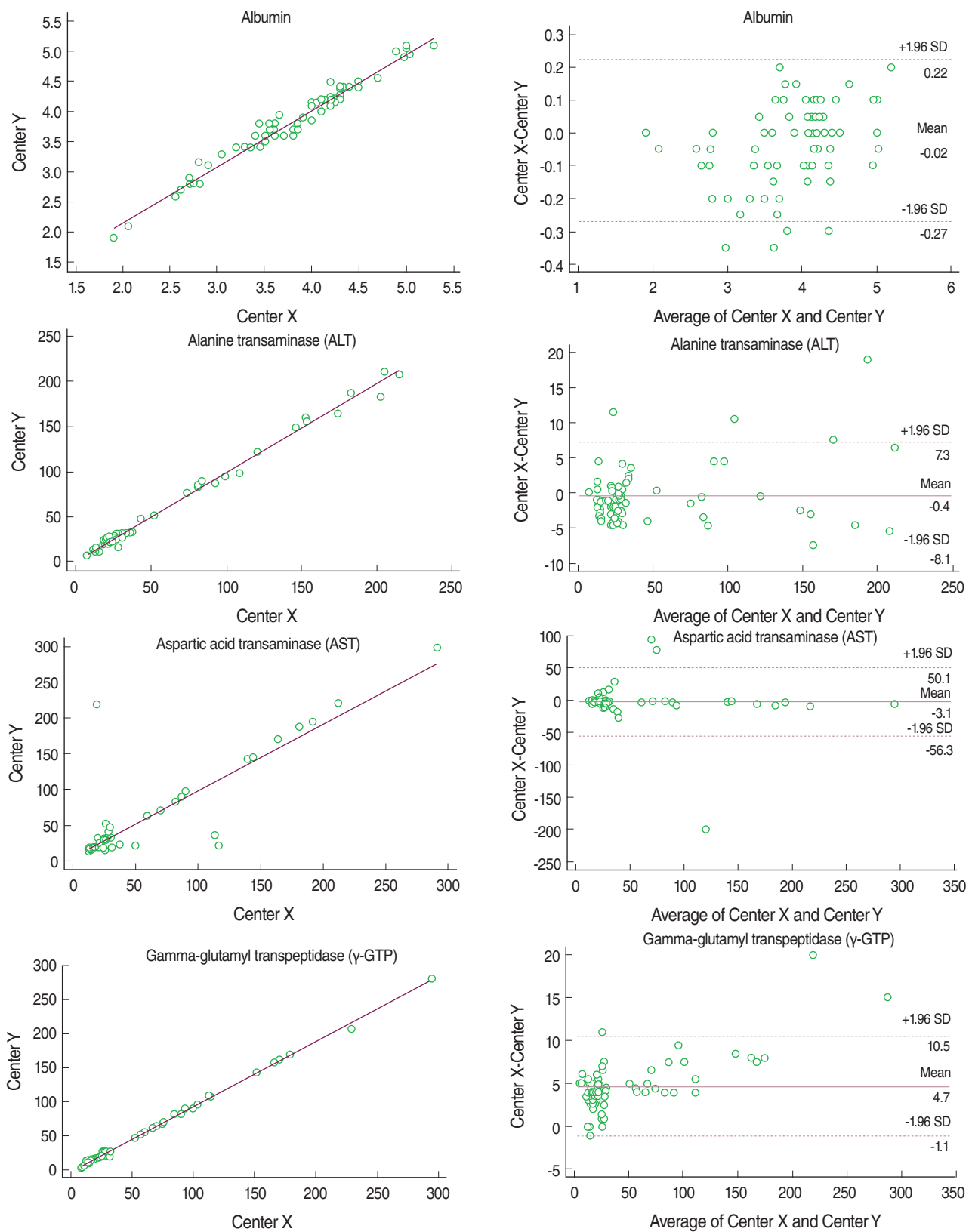
For intra-rater reliability at Center X, the mean differences ranged from 0.027 to 7.387. With the exception of AST (0.842), all correlation coefficients were greater than 0.997 ( $p < 0.001$ ). The ICC coefficients indicated excellent reliability that was greater than 0.836. Similarly, the mean differences at Center Y ranged from 0.018 to 3.125, all Pearson's correlation coefficients ranged from 0.989 to 0.999, and ICC coefficients ranged from 0.990 to 1.000, indicating almost perfect intra-rater reliability. Inter-laboratory reliability between Center X and Y was highly comparable. The means of the absolute values of the differences between the mean of the 40 duplicated test values at Centers X and Y ranged from 0.095 to 9.737. All correlation coefficients were reliable. Though AST showed the lowest Pearson's correlation coefficient and an ICC coefficient equal to the intra-rater

**Table 2.** Intra-rater reliability at each center and inter-laboratory reliability between the two test values at Centers X and Y

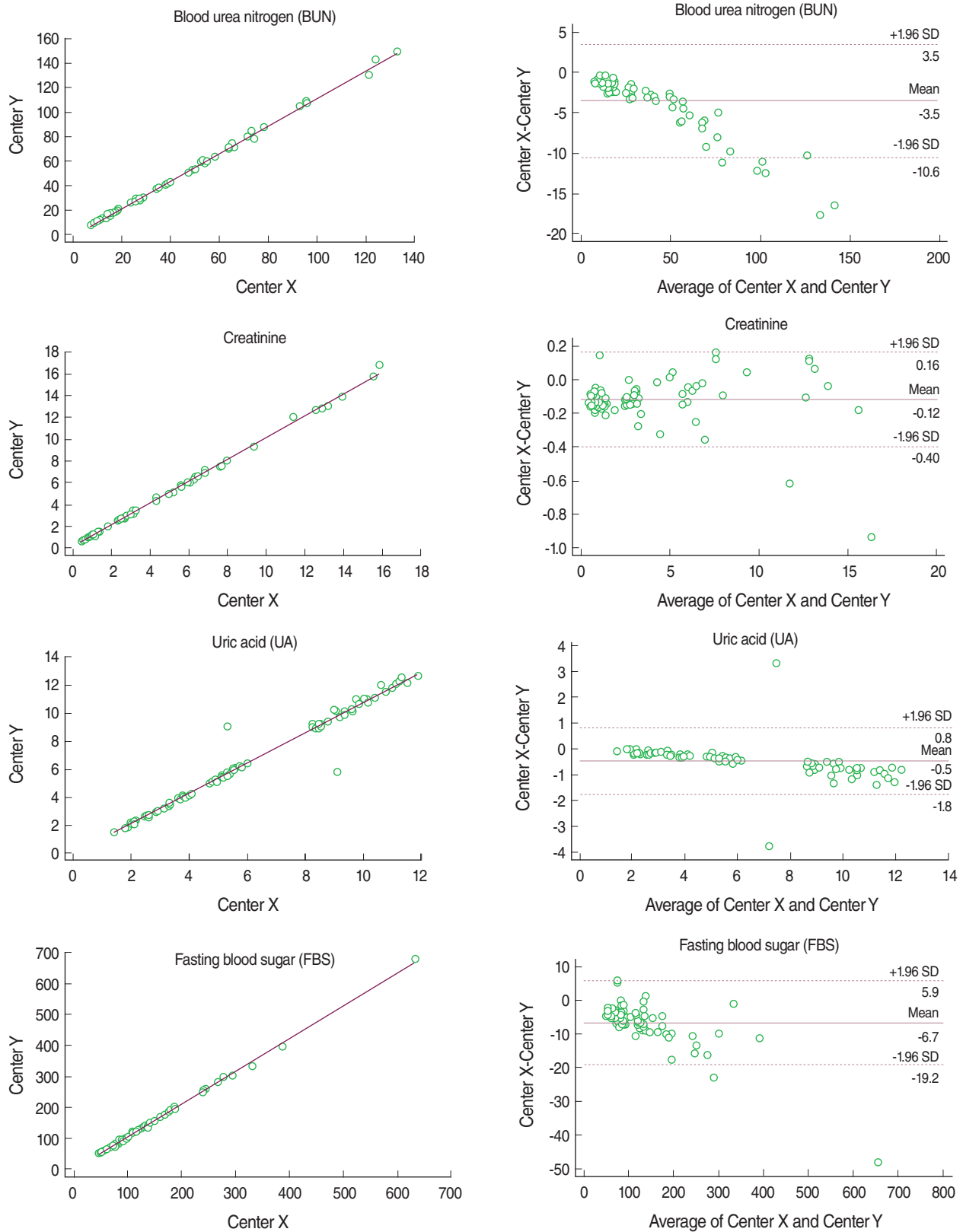
	Intra-rater reliability						Inter-laboratory reliability			
	Absolute value of mean (CV) of difference in two duplicated test values <sup>1</sup> at Center X			Absolute value of mean (CV) of difference in two duplicated test values <sup>1</sup> at Center Y			Absolute value of mean (CV) of difference between the two test values of Center X and Y <sup>2</sup>			
	Mean (CV) of difference	Correlation coefficient	ICC coefficient (95% CI)	Mean (CV) of difference	Correlation coefficient	ICC coefficient (95% CI)	Mean (CV) of difference	Correlation coefficient	ICC coefficient (95% CI)	Kappa coefficient (95% CI)
Albumin	0.027 (1.85)	0.997	0.997 (0.996-0.998)	0.018 (2.22)	0.998	0.998 (0.997-0.999)	0.095 (0.84)	0.986	0.984 (0.975-0.990)	0.962 (0.888-1.000)
ALT	0.437 (1.30)	0.999	1.000 (1.000-1.000)	2.173 (1.55)	0.996	0.997 (0.995-0.998)	2.693 (1.06)	0.997	0.997 (0.995-0.998)	1.000 (1.000-1.000)
AST	7.387 (4.12)	0.842	0.836 (0.755-0.891)	3.125 (2.34)	0.989	0.990 (0.984-0.993)	9.168 (2.80)	0.872	0.871 (0.807-0.915)	0.754 (0.567-0.941)
$\gamma$ -GTP	0.325 (1.75)	0.999	1.000 (1.000-1.000)	1.050 (1.09)	0.999	1.000 (0.999-1.000)	4.700 (0.62)	0.999	0.994 (0.796-0.999)	0.786 (0.637-0.934)
BUN	0.175 (2.34)	0.999	1.000 (1.000-1.000)	0.700 (1.17)	0.999	0.999 (0.999-1.000)	3.532 (1.02)	0.999	0.987 (0.901-0.996)	0.975 (0.925-1.000)
Creatinine	0.028 (1.79)	0.999	1.000 (1.000-1.000)	0.031 (1.61)	0.999	1.000 (1.000-1.000)	0.138 (0.87)	0.999	0.999 (0.995-1.000)	0.873 (0.765-0.981)
Uric acid	0.085 (2.59)	0.997	0.997 (0.996-0.998)	0.038 (1.84)	0.999	1.000 (1.000-1.000)	0.562 (1.03)	0.984	0.970 (0.893-0.987)	0.925 (0.842-1.000)
FBS	1.262 (1.42)	0.999	1.000 (1.000-1.000)	0.725 (1.53)	0.999	1.000 (1.000-1.000)	6.968 (0.87)	0.999	0.995 (0.952-0.998)	0.871 (0.761-0.981)
Cholesterol	2.162 (0.86)	0.998	0.998 (0.997-0.999)	2.587 (2.04)	0.992	0.992 (0.988-0.995)	5.850 (0.75)	0.992	0.988 (0.946-0.995)	0.876 (0.771-0.980)
TG	1.000 (1.03)	0.999	1.000 (1.000-1.000)	1.500 (2.43)	0.999	0.999 (0.998-0.999)	9.737 (0.65)	0.995	0.991 (0.914-0.997)	0.974 (0.922-1.000)

ALT, alanine transaminase; AST, aspartic acid transaminase;  $\gamma$ -GTP, gamma-glutamyl transpeptidase; BUN, blood urea nitrogen; UA, uric acid; FBS, fasting blood sugar; TG, triglyceride; CV, coefficient of variation; ICC, intra-class correlation.

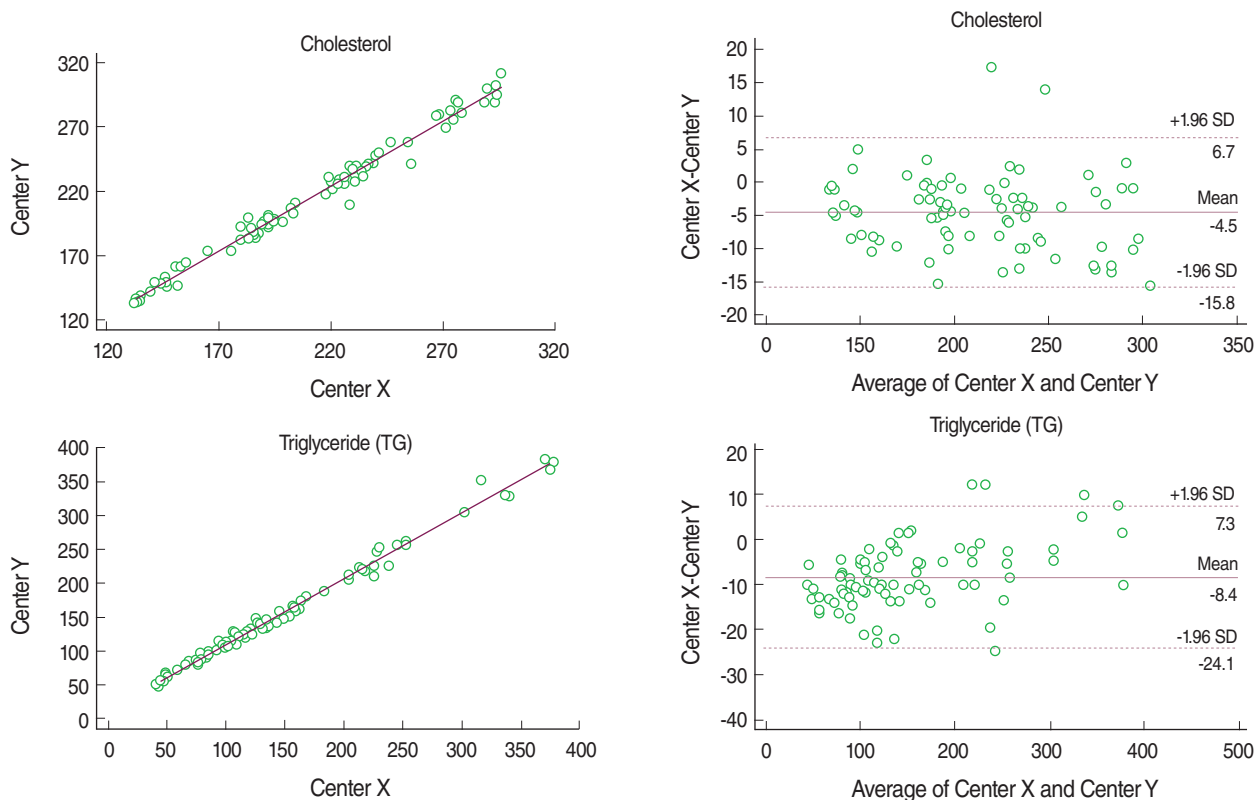
<sup>1</sup>Calculated as the absolute value of the difference in the mean of two duplicated test values at each center ( $n = 80$ ); <sup>2</sup>Calculated as the absolute value of the difference in the mean of two duplicated test values at Center X ( $n = 80$ ) and the mean of two duplicated test values at Center Y ( $n = 80$ ).



**Figure 1.** Scatter plots and Bland-Altman plots for each blood test item (*Continued to the next page*). SD, standard deviation.



**Figure 1.** (Continued from the previous page) Scatter plots and Bland-Altman plots for each blood test item. SD, standard deviation.



**Figure 1.** (Continued from the previous page) Scatter plots and Bland-Altman plots for each blood test item. SD, standard deviation.

**Table 3.** R-square values (sum of residuals)<sup>1</sup> on the fitting of linear regression models<sup>2</sup> using internal and external quality control data

	KoGES trial <sup>3</sup>	External QC data by KSLM <sup>4</sup>			Range of slope difference
		Year 2007	Year 2006	Year 2005	
Albumin	0.97	0.95	0.97	0.95	0.050-0.133
ALT	0.99	0.99	0.98	0.99	0.001-0.043
AST	0.70	0.99	1.00	1.00	0.011-0.206
γ-GTP	1.00	1.00	1.00	0.99	0.022-0.083
BUN	1.00	1.00	1.00	0.99	0.077-0.158
Creatinine	1.00	1.00	1.00	0.99	0.081-0.125
Uric acid	0.97	0.99	1.00	0.97	0.009-0.115
FBS	1.00	1.00	1.00	0.98	0.009-0.078
Cholesterol	0.98	0.99	1.00	0.96	0.007-0.030
TG	0.99	0.99	Not fitted <sup>5</sup>	Not fitted <sup>5</sup>	NA

ALT, alanine transaminase; AST, aspartic acid transaminase; γ-GTP, gamma-glutamyl transpeptidase; BUN, blood urea nitrogen; UA, uric acid; FBS, fasting blood sugar; TG, triglyceride; KoGES, Korean Genome and Epidemiology Study; KSLM, Korean Society for Laboratory Medicine; QC, quality control; NA, not applicable.

<sup>1</sup>Sums of residuals were all zero; <sup>2</sup>Equation forms are represented by (value of Center Y) =  $\alpha$  (intercept) +  $\beta$  (slope) × (value of Center X).

<sup>3</sup>The quadruplicated standardized samples from 40 samples; <sup>4</sup>Three standardized samples were tested for external QC four times a year by the KSLM; <sup>5</sup>Different QC methods applied in the two laboratories in 2005 and 2006.

reliability, the values remained good at 0.872 and 0.871, respectively. The kappa statistics showed attenuated reliability but all values were over 0.75 with substantial reliability (Table 2). Bland-Altman plots showed that most absolute differences between Center X and Y were within the 95% limit of agreement (Figure 1).

Using the standardized serological samples, statistical compensation models using linear regression analysis were constructed to yield a mathematical relationship between the results of the two laboratories. With the exception of AST ( $R^2=0.70$ ), all regression equations presented strong correlation coefficients ( $R^2>0.97$ ). The regression equations for the three years, 2005, 2006, and 2007, were estimated using external QC data acquired from the KSLM. All correlation coefficient values were greater than 0.95 with the exception of TG, which used a different QC method in the two laboratories in 2005 and 2006. All the results of the standardized serological samples and external QC data were consistent in the regression models (Table 3).

In the residual analysis, the sums of residuals were all zero and for those residual plots that presented residuals between the actual y-values and the predicted values all residuals were randomly distributed around zero (data not shown).

## DISCUSSION

This study evaluated the possibility of integrating serological parameters from different laboratories participating in the KoGES via a statistical adjustment algorithm with good intra- and inter-laboratory reliability. Our results indicated that the two laboratories had excellent intra- (correlation coefficient  $>0.84$ , ICC  $>0.83$ ) and inter-laboratory reliability (correlation coefficient  $>0.87$ , ICC  $>0.87$ , and kappa  $>0.75$ ) for ten chemical test values. Moreover, linear regression analysis to compensate for the discrepancy in test values between the two centers gave excellent R-square values and zero sums of residuals.

Poorly controlled data can lead to significantly biased results [11]. Given the participation of two laboratories in the KoGES, the issue of quality control should be carefully addressed. Likewise, a strategy for data integration should be established based on the reliability within and between the laboratories. The present study indicated that regression equations with higher R-square values can compensate for the potential discrepancies caused by the use of different laboratories. Although the absolute values differed slightly, if intra- and inter-laboratory reliability can be assured, data integration may be successfully conducted using statistical compensation models. In terms of AST,  $\gamma$ -GTP and TG with relatively unstable results or insufficient external QC data, further replication studies focused on clinical features and test methodology are required.

The issue of multiple laboratories is relevant not only for the KoGES but also for other large cohorts. Ideally, a single laboratory that passes a strict QC system should conduct all clinical tests according to accurate and standardized methods. However, this is logistically difficult in reality. The most reasonable alternatives are 1) choosing reliable laboratories, 2) developing a standardized protocol for clinical tests, 3) conducting and monitoring regular QC, and 4) integrating the KoGES database after statistical adjustment. For statistical adjustment, the regression analysis method can be used to check intra- and inter-rater reliability.

Using the KoGES data, this study aimed to determine the possibility of integrating serological parameters produced from different laboratories via statistical compensation models. Our results indicate that the ten blood chemical tests analyzed at the two laboratories can be integrated through statistical adjustments using regression equations. The existing external QC data should be used to correct the discrepancies in the other biochemical tests and complete blood cell counts, or additional tri-

als with the standardized serological samples should be conducted before data integration into the KoGES database.

## ACKNOWLEDGEMENT

This study was supported by a research grant from the Korea Centers for Disease Control and Prevention (2007-E71010-00).

## CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare for this study.

## REFERENCES

1. Khoury MJ, Beaty TH, Cohen BH. Fundamentals of genetic epidemiology. New York: Oxford University Press; 1993.
2. Korea Centers for Disease Control and Prevention. Korean genome and epidemiology study (KoGES). [cited 2008 Nov 18]; Available from: [http://cgs.cdc.go.kr/koges/a\\_a\\_a.jsp](http://cgs.cdc.go.kr/koges/a_a_a.jsp).
3. Schulze MB, Kroke A, Saracci R, Boeing H. The effect of differences in measurement procedure on the comparability of blood pressure estimates in multi-centre studies. *Blood Press Monit* 2002; 7: 95-104.
4. Iida M, Sato S, Nakamura M. Standardization of laboratory test in the JPHC study. Japan Public Health Center-based Prospective Study on Cancer and Cardiovascular Diseases. *J Epidemiol* 2001; 11: S81-S86.
5. McGuinness C, Secombe DW, Frohlich JJ, Ehnholm C, Sundvall J, Steiner G. Laboratory standardization of a large international clinical trial: the DAIS experience. DAIS Project Group. *Diabetes Atherosclerosis Intervention Study. Clin Biochem* 2000; 33: 15-24.
6. Whitney CW, Lind BK, Wahl PW. Quality assurance and quality control in longitudinal studies. *Epidemiol Rev* 1998; 20: 71-80.
7. Krouwer JS, editor. Method comparison and bias estimation using patient samples: approved guidelines. Wayne, PA: National Committee for Clinical Laboratory Standards; 2002.
8. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307-310.
9. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 1973; 33: 613-619.
10. Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991.
11. Desmet M, Linder SN. The challenge of auditing clinical laboratories. *Drug Inf J* 1997; 31: 197-205.